Reviewer#2

This paper by Sauvage et al., presents a system (SOFT-IO) based on the extensive use of FLEXPART dispersion model (coupled with different inventories of anthropogenic and fire emissions), created to analyse and attribute the variability of atmospheric composition observed along a huge number of observations by the IAGOS-MOZAIC programme. Even if, in this current configuration, the system is able to simulate only CO variability, it is valuable for the interpretation of this important long-term data base. From my understanding, the SOFT-IO outputs will be easily accessible to external users and thus they represent a potentially powerful tool for a number of applications. Since the system is based on a pre-computed data-set of air-mass transport simulation by FLEXPART model, it is possible to couple it with other emission inventories besides those used in this work. As a personal comment, it would be really great if this system will be made available also for other observation systems (e.g. WMO/GAW stations). Other than presenting SOFT-IO tool, the paper also provides an assessment of its performance in correcting reproducing the variability of observed CO due to anthropogenic and fire emissions over differentWorld regions (where the IAGOS-MOZAIC programme is/was active) also discussing (by mean of case study analysis, and sensitivity studies) the dependency of SOFT-IO results as a function of different parameters (i.e. different input meteorological data-set, different emission inventories, different scheme for pyroconvection).
By discussing the differences between SOFT-IO simulations and observations, the paper also provides information about the accuracy of different emission inventories or pyro-convection schema.
The paper is clear and very well written and I strongly recommend publication after that some points (most of them, minor) are considered.

We would like to thank Reviewer#2 for her/his comments and suggestions that will improve our manuscript. We clarified all the points raised by reviewer#2 and answered her/his different remarks or comments in blue in this document and in the revised manuscript.

However, I have to stress (this is my only major concern) that the scientific significance of the SOFT-IO simulations are only limited discussed. As an instance, the authors provided very interesting longterm time series of CO over different regions of the World but without giving any comments or indications about differences among regions, about the existence/attribution of long-term trends (both in observations and simulations) , about seasonal variability or SOFT-IO agreement with other data-sets apart MOZAIC.

Rev#2 is true that there is limited discussion of the scientific significance of SOFT-IO simulations. This choice is deliberate. Indeed as stated in the "Introduction" section (lines 86-90 of the submitted manuscript), the goal of the paper is to present and validate SOFT-IO as well as the CO ancillary products calculated with SOFT-IO, for the IAGOS database and the IAGOS users.
Ancillary products calculated with SOFT-IO will then be implemented in the IAGOS database, so that further scientific interpretations of the IAGOS data using SOFT-IO will follow in future papers realized by IAGOS database users.
For instance, long-term CO series have been first analyzed in a recent study of Cohen et al. (2017, https://doi.org/10.5194/acp-2017-778). Our study just aim to evaluate SOFT-IO in terms of long-term series reproducibility (Figure 11). The use of SOFT-IO to comment the existence/attribution, or to give indication about the differences in regional trends will be done in further studies, out of the scope of this paper. However, this is definitely our

objective, to go further for example regarding the CO trends analysis than the recent Cohen et al., 2017 work.
Moreover, as stated in the introduction, the main goal of SOFT-IO is to provide ancillary data that should help the IAGOS users interpreting the IAGOS database. SOFT-IO source code will be available as soon as the paper would be accepted, so that everybody could use it on other data-sets as suggested by Ref#2. We encourage external users to apply SOFT-IO to other dataset, such as ground based CO measurements, or for CO aircraft campaigns. However it is out of the scope of this study to evaluate the model to other dataset but IAGOS. Indeed IAGOS represents to our knowledge the densest in-situ measurements CO dataset, and it will be easier to apply SOFT-IO to other in situ CO datasets.

We modified the following sentence lines 86-90 for more clarity:
"*The goal is to provide the scientific community with added value products that will help them analyzing and interpreting the large number of IAGOS measurements. The methodology is focused on the development of a scientific tool (SOFT-IO version 1.0) based on FLEXPART particle dispersion model, that simulates the contributions of anthropogenic and biomass burning emissions for IAGOS CO measurements. This tool, which has the benefit to be adaptable to multiple emission inventories without re-running FLEXPART simulations, is described and then evaluated in the present study with the large data-sets of IAGOS CO measurements. SOFT-IO could be in the future easily adapted and used to analyze other datasets of trace gas measurements such as from ground based observations, sondes, aircraft campaigns or satellite observations.*"

In the same way, possible
limitations/inaccuracy of the considered emission inventories (which have been pointed out by the authors) must be better addressed/discussed also in view of their extensive use in air-quality or climate studies.
In the same way, we deliberately did not discuss the limitations and accuracy of the emission inventories. This is out of the scope of the paper. SOFT-IO could be in a future a useful tool to investigate emission inventories limitations or accuracy by the scientific community in charge of developing emission inventories, or investigating air quality or climate studies. The present paper only aims to present the SOFT-IO tool developed to help IAGOS users interpreting a large database such as the IAGOS one, to evaluate the tool against IAGOS data. Therefore, we provide these CO contributions to the IAGOS users as added-value products.

Finally, I visited the IAGOS web site but I was not able to find SOFT-IO output. Probably, they are still not available to external users: : :
Rev#2 is true, we believe that our validation paper should be accepted to make the code and the data available for external users.

Minor/technical points

1) Figure 2: it seems that for boreal fires (with FRP > 10 Tjday)
the injection fraction decrease with height along the first atmospheric layers (up to 2000 m). It is correct? This is the effect of atmospheric vertical mixing/stability?
For boreal fires (> 10 and < 100TJ/day; > 100TJ/day), the injection fraction decreases with height higher than 3000m. Indeed this is the effect of atmospheric vertical mixing, as calculated by the PRMv2 model.

2) In general the figure should be better arranged. I would recommend the authors to reshape the plots so that each full figure (often composed by several plates) can be showed in a single page. This would help the reader also in comparing the results of the sensitivity tests

We have arranged the figures so that they are on a single page.

3) Table 3: please provide some statistical indications to provide quantitative indication about the agreement for the two inventories (e.g. by providing average CO values for observations and simulations, mean bias, timing of the detected peak, std. dev..)

We have added statistical information for Table 3 and Table 4 (see below)

4) Pag 6. To me is not clear how the injection profile is defined: : :please clarify it.

The injection profile is defined according to three methodologies, as explained page 7 lines 225-252, the DENTENER, the MIXED or the APT one.

In order to clarify, we add the following sentence lines 217-218:

" and defined according to three different approaches (DENTENER, MIXED or APT) described in the next paragraph"

5) Pag. 10. It's not clear why you claimed that only 2/3 of peaks are simulated by EDGAR. In my opinion, all the peaks are simulated by EGARD run indeed

All the peaks are simulated by EDGAR, but only 2/3 of the peaks intensity is reproduced using EDGAR.

We will rephrase lines 374-375

"*Only 2/3 of the observed enhancements are simulated using EDGARv4.2, except for plume 1 with better results*" with "*Using EDGARv4.2, only 2/3 of the observed CO enhancements intensity is reproduced, except for plume #1 with better intensity results*"

6) Fig. 11, line 413. Thus the incorrect quantification of the bottom part of the peak by the ICARTT run can be attribute to not perfect transport/mixing by FLEXPART? Please comment, on that.

It seems that Rev#2 refers here to Fig.7a. In this case it is hard to explain why the bottom part of the peak is not represented as well as the upper part, either by ICARTT, GFED or GFAS. It could indeed be related to transport processes in FLEXPART, but also in the ECMWF analyses or in the vertical profiles injection.

7) Pag 12, Figure 9: it can be interesting also to separate the plumes attributed to fires from these due to anthropogenic emissions .

Ideally this could be interesting. But this is not possible to realize. Indeed all the plumes are influenced both by biomass burning and anthropogenic emissions, as we can see on the case studies displayed on Figures 5 to 8. In order to do that we should define subjective criteria to attribute a plume to either biomass burning or anthropogenic emissions. This is out of the scope of this study.

8) Pag 13, line 493: I would say that for North Asia UT discrepancies varied from -100 to + 200 ppb and for South Asia LT from – 50 to +100 ppb.

We have modified line 493 with Rev#2 suggestion:

"*North Asia UT discrepancies varies from -100 ppb to +200 ppb and from -50 ppb to +100 ppb for South Asia LT.* "

9) Pag 14, line 516: the possible misrepresentation of anthropogenic emissions after 2009 is a point of great importance that deserve more discussion. The overestimation

in the MT appeared to be more and more relevant over NAM than EU. Please comment.
Rev#2 is true. As stated in Stein et al., 2004, the largest near-surface CO bias are found over Europe in January.
We have added the following lines 519:
"This suggests misrepresentation of anthropogenic emissions in Europe after the year 2009. Indeed Stein et al., (2014) suggested the lower near-surface CO bias was found in Europe in relation with possible under estimation of traffic emissions in the inventories."

It is also true that the overestimation in the MT appears higher over NAM rather than over EU. This could be related to two causes:
- Less measurements in the MT over NAM than over EU
- Greater proximity of the NAM MT to summer sources, such as boreal fires, that could explain the higher overestimation particularly in this season.

We add the following lines 519-529
"In the middle troposphere (2-8 km), the CO plumes are systematically overestimated by SOFT-IO by 50% to 100% compared to the observations, with larger standard deviation and higher overestimation over NAm. This might be related to different reasons:

- *the chosen methodology of the CO plume enhancements detection for those altitudes (described in Sect. 3.4), which may lead to a large number of plumes with small CO enhancements, which are difficult to simulate. This could be due to the difficulty in defining a realistic CO background in the middle troposphere.*

- *the source-receptor transport which may be more difficult to simulate between 2-8 km than in the LT where receptors are close to sources; or than in the UT where most of the plumes are related to convection detrainment better represented in the models than MT detrainment which might be less intense.*

- *The frequency of the IAGOS observations which is lower in the LT and in the MT than in the UT.*

- *Higher overestimation over NAm MT than Eur MT could be first related to lower frequency of measurements in the NAm. Moreover overestimation is greater during summer when NAm MT is closer to summer sources such as boreal fires, while Eur MT is related to CO air masses more diluted with background air during transatlantic transport. "*

10) Pag 15, line 559: I would not say that EDGAR performed better that MACC inventory for CAS_MT and NAS_UT: are these differences really significant?
Indeed results are better using EDGAR for specific regions. Ref#2 is right. Differences are not statistically significant for NAs_UT, but they are for CAs_MT (almost 50% difference between the two simulations with the two inventories).
We rephrase line 559 with the following *"Regionally, however, results with EDGARv4.2 can be better by almost 50%, such as over South Asia LT and MT, Central Asia LT and MT"*

Tables modifications requested by Rev#2:

Flight	IAGOS anomaly	IAGOS std	MACCity anomaly	MACCity std	EDGAR anomaly	EDGAR std	Anomaly altitude
10 March 2002 Frankfurt – Denver	16.8	8.7	20.2	6.9	12.8	5.1	UT
27 November 2002 Dallas – Frankfurt	28.0	8.6	20.0	8.0	16.4	7.4	UT
19 July 2005 München - Hong Kong	130.1	97.8	45.8	9.7	34.6	7.7	PBL
22 October 2005 München - Hong Kong	157.9	105.1	170.7	109.8	103.9	62.0	PBL

Table 3. Summary of the averaged observed and simulated anomaly and corresponding averaged standard deviation (std) (in ppb) determined for representing anthropogenic emissions for different case studies (using GFAS v1.2 for biomass burning emissions). Altitude of the anomaly is indicated: boundary layer (PBL); middle troposphere (MT); upper troposphere (UT)

Flight	IAGOS anomaly	IAGOS std	GFAS v1.2 anomaly	GFAS v1.2 std	GFED4 anomaly	GFED4 std	ICARTT anomaly	ICARTT std	Anomaly altitude
29 June 2004 Caracas - Frankfurt	32.6	33.2	44.4	2.4	43.0	2.3	43.6	2.4	PBL
30 June 2004 Frankfurt - Washington	52.5	34.0	36.6	9.1	25.4	6.6	23.5	5.9	MT
22 July 2004 Frankfurt - Atlanta	87.0	35.0	42.8	17.6	45.8	18.9	39.7	15.7	MT
22 July 2004 Douala - Paris	117.1	24.2	43.5	20.0	55.0	27.2	72.4	42.3	MT
23 July 2004 Frankfurt - Atlanta	78.9	45.4	34.7	22.4	45.3	32.8	46.0	35.9	MT
30 July 2008 Windhoek - Frankfurt	72.9	41.9	33.0	19.2	42.8	26.0	N/A	N/A	UT
31 July 2008 Frankfurt - Windhoek	38.3	32.0	28.1	10.8	34.0	12.8	N/A	N/A	UT

Table 4. Summary of the averaged observed and simulated anomaly and corresponding averaged standard deviation (std) (in ppb) determined for representing biomass burning emissions for different case studies (using MACCity for anthropogenic emissions). Altitude of the anomaly is indicated: boundary layer (PBL); middle troposphere (MT); upper troposphere (UT). Note that the ICARTT inventory is only available for summer 2004.